



---

# IHS National Data Warehouse Project Plan (Draft)

**IHS Data Quality Action Team  
Last Revised June 21, 2001  
Version 1.1**

## Table of Contents

<b>1.</b>	<b>EXECUTIVE SUMMARY .....</b>	<b>1</b>
1.1.	Goals of the Data Warehouse .....	1
1.2.	IBM Assessment .....	2
1.3.	Conceptual Model of Data Warehouse .....	2
1.4.	Overview of Project Schedule .....	2
1.4.1.	Phase I – Pilot .....	2
1.4.2.	Phase II – Full Implementation of Patient Registration and Medical Records Subject Group .....	4
1.4.3.	Phase III – Full Implementation of Remaining Subject Groups..	5
1.5.	Summary.....	6
<b>2.</b>	<b>COORDINATION WITH OTHER DATA EFFORTS .....</b>	<b>6</b>
<b>3.</b>	<b>PROJECT TEAM .....</b>	<b>7</b>
	<b>IBM ASSESSMENT .....</b>	<b>APPENDIX A</b>
	<b>DETAILED PROJECT SCHEDULE .....</b>	<b>APPENDIX B</b>

DRAFT

# National Data Warehouse Project

## 1. EXECUTIVE SUMMARY

The IHS Data Quality Action (DQA) Team, a team appointed by senior IHS management, has been tasked with implementing changes to provide accurate, measurable, and timely improvements to national reporting systems. As part of this and consistent with recent recommendations from the Data Management Committee, the DQA Team has begun to design and implement an IHS national data warehouse (DW) and associated data marts. The DW will be developed independently of the current NPIRS production database and ultimately will be the primary IHS national database, storing information across and providing service to the entire I/T/U spectrum.

### 1.1. Goals of the Data Warehouse

The goals of the DW are to:

- ❑ Implement a true relational database design.
- ❑ Provide web-based access to quality, accurate data to IHS' customers in a timely, user-friendly manner while at the same time protecting the privacy of the data.
- ❑ Increase customer confidence in the data.
- ❑ Decrease the turnaround time required for processing exported data and making the data available to IHS' customers, whether it is a data mart, report, view, or a database query.
- ❑ Develop specialized data marts to be used by programs such as NPIRS for user population and workload reporting, Epidemiology for the study and analysis of patient health care data as it relates to the cause and control of epidemics, ORYX for facility accreditation, and others.
- ❑ Provide information to the owners of the data immediately after receiving the data, such as timeliness of the data, potential problems with the data, etc.
- ❑ Integrate the many facets of IHS data, including patient registration and medical records data, provider and employee data, facility and equipment asset data, financial data, and product and services data. These functional areas are referred to as DW Subject Groups.



DRAFT

## 1.2. IBM Assessment

IHS recently contracted with IBM to perform an assessment of the current NPIRS database structures and processes and to develop a strategic approach for planning and implementing a DW. IBM has completed its draft assessment, which is currently being reviewed by the DQA Team and senior IHS management. Once officially accepted by IHS, the information presented will provide the first steps toward implementing the DW.

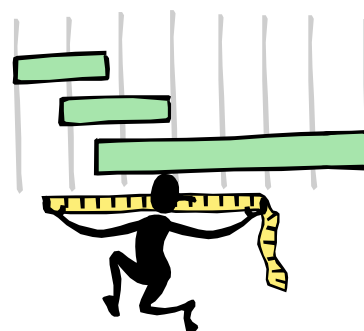
This assessment will be included as Appendix A in this document once it has been reviewed and accepted by IHS. It will also be posted on the DQA Team's website.

## 1.3. Conceptual Model of Data Warehouse

The graphic on the next page illustrates a conceptual model of the IHS National DW.

## 1.4. Overview of Project Schedule

This project will be sub-divided into three phases, with the first phase expected to begin June 25, 2001. A detailed project schedule will be developed in MS Project and included in Appendix B. It will also be made available on the DQA Team's website.



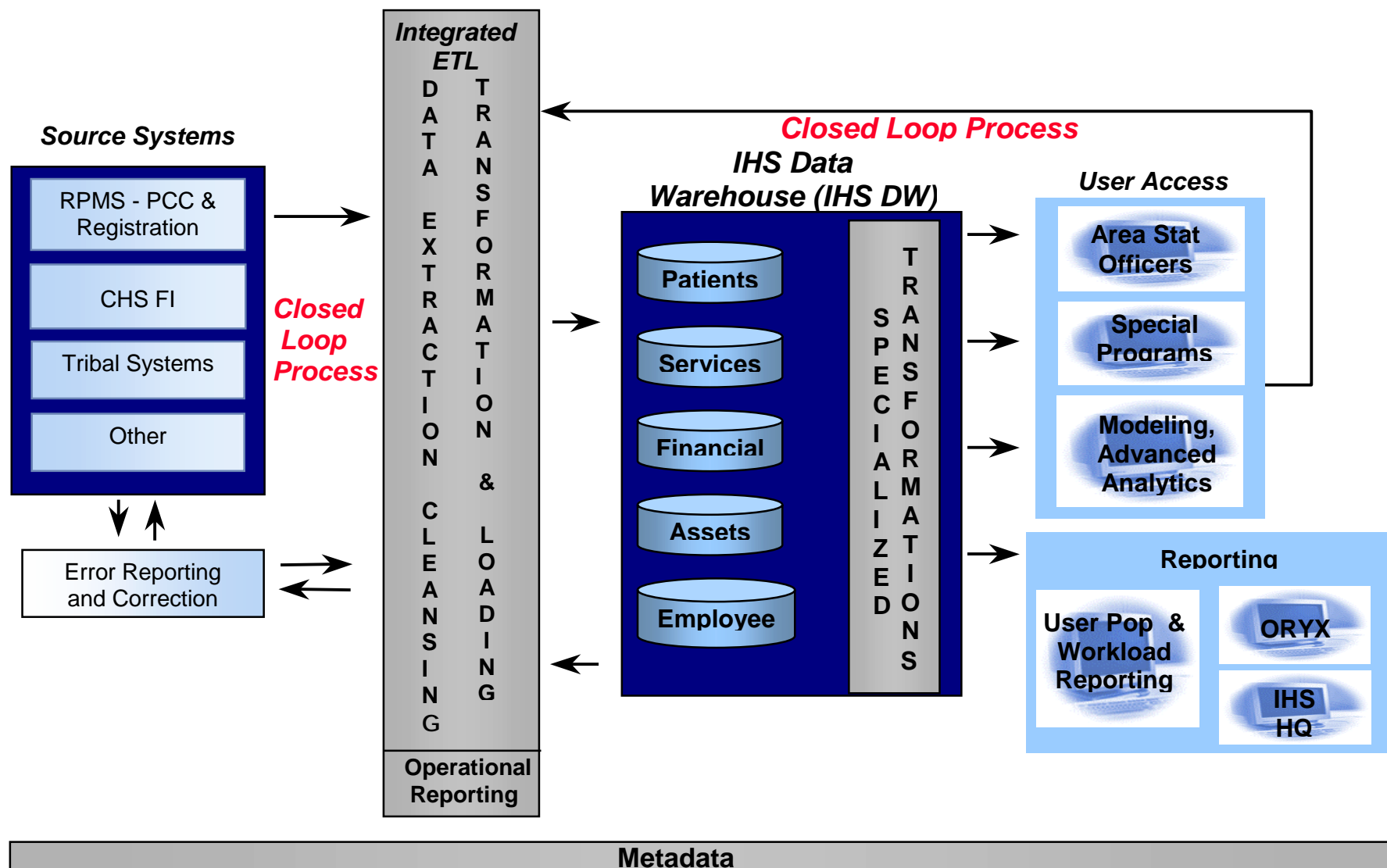
### 1.4.1. Phase I – Pilot

The first phase is to develop a pilot DW and is expected to last approximately three to four months. Three IHS Areas - Albuquerque, Nashville, and Phoenix - have been asked to participate in the pilot phase. The key events that will occur in this phase are listed below. NOTE: Due to time constraints and limited funding resources, no data marts will be created during this phase.

- ❑ Develop and implement the logical and physical data models for the patient registration and medical records (i.e., visit) data, which includes the creation of the databases. This will provide a 90-95% solution to the data modeling requirements associated with patient registration and medical records data.
- ❑ There will be two separate databases created, one for use by IBM and for SAS (see second bullet on page four).

DRAFT

## Conceptual Model of an IHS National Data Warehouse



DRAFT

- ❑ Using enhanced export programs, Areas re-export their data, including both patient registration<sup>1</sup> and FY97 to present medical records data. This will also include data from tribally-operated facilities that send their data directly to NPIRS, such as Penobscot in Nashville Area.
- ❑ IBM and SAS demonstrate their applications for extracting, transforming, and loading (ETL) of the re-exported data on separate databases. Simply put, ETL is the processing of the raw data files in which the data are stored in the DW.
- ❑ Design and run Workload and User Population reports which are generated from the DW. Compare the reports produced by IBM with the SAS reports, as well as compare both of these reports with those produced by NPIRS. Research and correct any problem areas.
- ❑ Make reports available to HQ and Area Statistical Officers. Work together to confirm the data warehouse and reporting functionality are working correctly.
- ❑ Areas transmit a separate export of data that will include modified as well as new records so the functionality of updating existing records can be tested in the DW.
- ❑ IHS selects the application that is the most cost effective and efficient and begins using it for managing the ETL processes in the DW.
- ❑ Develop procedures and data structures to handle metadata (i.e., documentation) associated with the DW.
- ❑ Identify lessons learned during this phase and changes that will need to be made during Phase II.
- ❑ Develop User Requirements for a complete implementation of patient registration and medical records data in Phase II.

#### **1.4.2. Phase II – Full Implementation of Patient Registration and Medical Records Subject Group**

This phase will begin after completion of Phase I and is expected to be completed within four to five months. The key events that will occur in this phase are listed below.

---

<sup>1</sup> IHS HQ will be sending a message to all Areas informing them to re-export all of their patient registration data to NPIRS. Currently there is a Statistical Officer's Workgroup that is developing the methodology to be used by all Areas for re-exporting their patient registration data. If during Phase I, the workgroup is unable to develop the methodology in time so the three Areas can use it for re-exporting, these Areas will most likely use the methodology used by Phoenix Area when it re-exported in 2000. Thus, when the methodology does become available, the three Areas would need to re-export their data again, using this new methodology. The DQA Team is going to communicate to the members of the workgroup the importance of having the methodology available when Phase I is ready for the re-exporting process.

DRAFT

- ❑ If necessary, revise the logical and physical data models.
- ❑ Completely populate the DW with patient registration and FY97 to present medical records data for all source systems. For the Areas that participated in Phase I, their patient registration and medical records data that was re-exported in Phase I would need to be supplemented with data that was exported after the re-export took place. Thus, these three Areas would only need to re-export that data. However, there are two exceptions to this. The first is explained in the footnote on page two of this document. The second is regarding the medical records data. If the PCC export is revised to include additional fields and/or to correct problems with existing fields, such as the Unique ID field, and is distributed, the three Areas would need to re-export their data using the new version of the PCC export.
- ❑ Design, test, and implement the initial data marts, such as Workload and User Population reporting, ORYX, Epidemiology, and GPRA reporting. This will include addressing and implementing access and security requirements to accommodate a wide variety of users.
- ❑ Run Workload and User Population reports which are generated from the Workload and User Population Reporting data mart. Compare reports with those produced by NPIRS.
- ❑ Make reports available to HQ and Area Statistical Officers. Work together to confirm the data mart and reporting functionality are working correctly.
- ❑ Assume responsibility as the IHS national repository for data. At that time, NPIRS will be transitioned to a data mart that is fed from the DW, and the current NPIRS database will be deactivated.

#### **1.4.3. Phase III – Full Implementation of Remaining Subject Groups**

This phase will begin after completion of Phase II. The duration of this phase is unknown at this time. The key events that will occur in this phase are listed below.

- ❑ Develop and implement the logical and physical data models, which includes creation of the databases for the remaining subject groups, as determined by IHS working in collaboration with Tribal and Urban programs. The remaining subject groups are:
  - Services
  - Financials
  - Assets (NECOP and STORES)
  - Employees

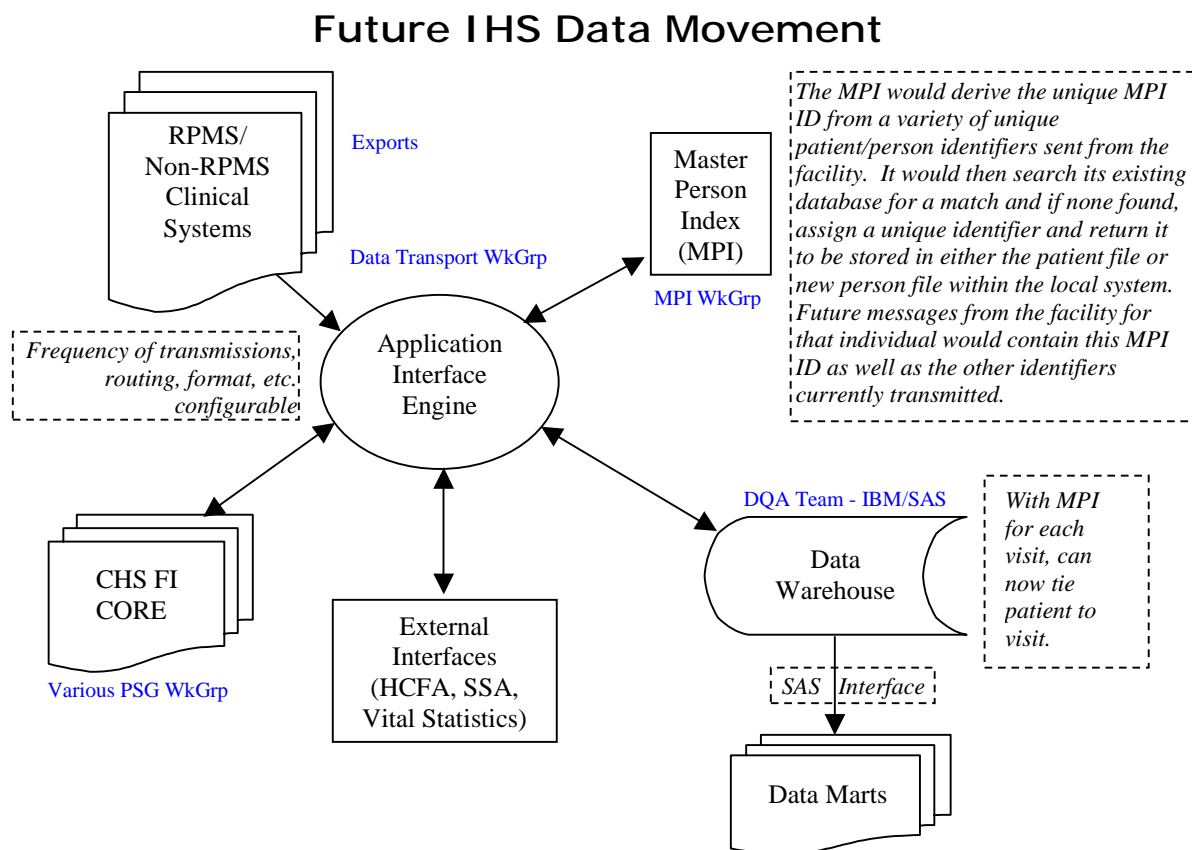
DRAFT

## 1.5. Summary

IHS recognizes there have been significant problems in the IHS national data repository (i.e., NPIRS), particularly the inability to produce accurate and timely User Population and Workload reports and lack of access to this data those who need it most. IHS is listening to its customers and acknowledges that these problems cannot, and will not, continue. After careful review, ITSC has decided to start fresh by implementing a new national data warehouse. If IHS is successful, and it will do everything in its power to ensure that it is, this data warehouse will allow for the integration of data across and provide service to the entire I/T/U spectrum, using the latest in warehousing technologies. And, of most importance, the data will be accurate, timely, secure, and yet easily accessible by those with proper authorizations, things that are critical to the mission of IHS.

## 2. COORDINATION WITH OTHER DATA EFFORTS

The National Data Warehouse project will be coordinated with the other data efforts that are currently taking place in IHS, such as the MPI and Data Transport workgroups. However, the implementation of the National Data Warehouse project is not dependent on completion of any or all of these other projects. These combined efforts will result in a revised data movement flow, as illustrated below.



DRAFT

### 3. PROJECT TEAM

The members of the National Data Warehouse implementation team are listed below:

- ❑ **Project Lead:** Rus Pittman, Director, ITSC
- ❑ Karen Carver, IHS Data Quality Action Team
- ❑ Joan Christy, IHS Data Quality Action Team
- ❑ Mike Danielson, Health System Administrator, Billings Area
- ❑ Mike Gomez, ORYX Program Manager
- ❑ Stan Griffith, Acting Lead, IHS Data Quality Action Team
- ❑ Joe Herrera, NPIRS Database Administrator (DBA)
- ❑ Stephanie Klepacki, IHS Data Quality Action Team
- ❑ Danny Macias, ORYX

These members will work closely with IBM and/or SAS personnel to communicate clearly IHS' business rules, data movement processes, etc., in order to produce an efficient and effective national data warehouse according to the project schedule.

DRAFT

# APPENDIX A

## IBM ASSESSMENT

## **APPENDIX B**

### **DETAILED PROJECT SCHEDULE**